段階的 多變量 線型回歸에 관하여

申 敏 雄*
金 周 成**

I. 序 論

단계적 선형회귀모형

$$Y=BX+E \tag{1-1}$$

를 생각하자. 여기서 Y는 관찰치들의 $p\times N$ 행열, B는 모회귀계수의 $p\times q$ 행열, X는 階數(rank) $q(<N)$인 고정된 변량들의 $q\times N$행열이고, E는 오차항의 $p\times N$행열이다. 지금 B와 X가 적당히 몇개의 행열로 분할되어진다고 하자.

즉 $B=(B_1\ B_2), X'=(X_1'\ X_2')$.

그러면 (1-1)은

$$Y=B_1X_1+B_2X_2+E \tag{1-2}$$

로 표시할 수 있다.

B_1과 B_2의 최소자승 추정량, 즉 $\hat{B}_1$와 $\hat{B}_2$는

$$\begin{bmatrix} X_1X_1' & X_1X_2' \\ X_2X_1' & X_2X_2' \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} Y' = \begin{bmatrix} \hat{B}_1' \\ \hat{B}_2' \end{bmatrix} \tag{1-3}$$

로 주어지는 반면에 이단계적 추정량 B_1과 B_2는

$$\begin{bmatrix} (X_1X_1')^{-1}X_1 \\ (X_2X_2')^{-1}X_2D \end{bmatrix} Y' = \begin{bmatrix} \bar{B}_1' \\ \bar{B}_2' \end{bmatrix} \tag{1-4}$$

단, $D=I-X_1'(X_1X_1')^{-1}X_1$

로 쉽게 구해진다.

$\hat{B}_1$와 $\hat{B}_2$는 B_1과 B_2의 불편추량치이지만, $\overline{B_1}$와 $\overline{B_2}$는 불편추정량이 아니다.

본 논문의 목적은 $\hat{B}_1$, $\hat{B}_2$와 $\overline{B_1}$, $\overline{B_2}$ 사이에 생기는 편차를 구하는 데 있다. 우리는 참고문헌 〔4〕와 다른 방법으로 구하였다.

* 충남대학교 문리과대학 계산통계학과
** 충남대학교 대학원 수학과

II. 模型과 偏倚

다변량 선형회귀 모형

$$Y=BX+E \tag{2-1}$$

를 생각하자. 즉

$$\begin{bmatrix} Y_{11} & \cdots & Y_{1N} \\ Y_{21} & \cdots & Y_{2N} \\ & \cdots\cdots & \\ Y_{p1} & \cdots & Y_{pN} \end{bmatrix} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1q} \\ \beta_{21} & \cdots & \beta_{2q} \\ & \cdots\cdots & \\ \beta_{p1} & \cdots & \beta_{pq} \end{bmatrix} \begin{bmatrix} X_{11} & \cdots & X_{1N} \\ X_{21} & \cdots & X_{2N} \\ & \cdots\cdots & \\ X_{q1} & \cdots & X_{qN} \end{bmatrix} + \begin{bmatrix} \epsilon_{11} & \cdots & \epsilon_{1N} \\ \epsilon_{21} & \cdots & \epsilon_{2N} \\ & \cdots\cdots & \\ \epsilon_{p1} & \cdots & \epsilon_{pN} \end{bmatrix}.$$

$B=(B_1\ B_2)$, $X'=(X_1'\ X_2')$라 하면

$$\begin{aligned} Y&=BX+E \\ &=[B_1\ B_2]\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}+E \\ &=B_1X_1+B_2X_2+E \end{aligned} \tag{2-2}$$

여기서

$$Y_i=(Y_{1i},\ Y_{2i},\cdots,Y_{pi})'$$

$$B_i=(\beta_{1i},\ \beta_{2i},\cdots,\beta_{pi})'$$

$$\epsilon_i=(\epsilon_{1i},\ \epsilon_{2i},\cdots,\epsilon_{pi})'\quad i=1,\cdots,n$$

로 놓으면

$$(Y_1,\cdots,Y_N)=(B_1,\cdots,B_q)\begin{bmatrix} X_{11} & \cdots & X_{1N} \\ & \cdots\cdots\cdots & \\ X_{q1} & \cdots & X_{qN} \end{bmatrix}+(\epsilon_1\cdots\epsilon_N) \tag{2-3}$$

양변의 행열을 전치하면, 즉

$$\begin{bmatrix} Y_1' \\ \vdots \\ Y_N' \end{bmatrix} = \begin{bmatrix} X_{11}\cdots X_{q1} \\ \cdots \\ X_{1N}\cdots X_{qN} \end{bmatrix} \begin{bmatrix} B_1' \\ \vdots \\ B_q' \end{bmatrix} + \begin{bmatrix} \epsilon_1' \\ \vdots \\ \epsilon_N' \end{bmatrix}$$

B_1과 B_2의 일단계 최소자승 추정량, 즉 $\hat{B}_1$와 $\hat{B}_2$는

$\hat{B}'=(XX')^{-1}XY'$ 이므로

$X'=(X_1'\ X_2')$, $\hat{B}'=\begin{bmatrix} \hat{B}_1' \\ \hat{B}_2' \end{bmatrix}$를 대입하여

$$\begin{aligned} \begin{bmatrix} \hat{B}_1' \\ \hat{B}_2' \end{bmatrix} &= \left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}(X_1'\ X_2')\right]^{-1}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}Y' \\ &=\begin{bmatrix} X_1X_1' & X_1X_2' \\ X_2X_1' & X_2X_2' \end{bmatrix}^{-1}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}Y' \end{aligned} \tag{2-4}$$

그리고 이단계 추정치, $\overline{B}_1'$와 $\overline{B}_2'$는 $Y'=X_1'B_1'+E_1'$ 이므로

$$\overline{B}_1'=(X_1X_1')^{-1}X_1Y'$$

가 된다.

다음, 잔차 $\tilde{Y}$는

$$\tilde{Y} = Y' - X_1'\overline{B_1}'$$
$$= Y' - X_1'(X_1X_1')^{-1}X_1Y'$$
$$= (I - X_1'(X_1X_1')^{-1}X_1)Y' \tag{2-5}$$

$\tilde{Y}$를 X_2에 관하여 회귀하면 회귀계수 $\overline{B}_2'$를 구할수 있다.

$$\overline{B}_2' = (X_2X_2')^{-1}X_2\tilde{Y}'$$
$$= (X_2X_2')^{-1}X_2(I - X_1'(X_1X_1')^{-1}X_1)Y'$$

그러므로

$$\begin{bmatrix} \overline{B}_1' \\ \overline{B}_2' \end{bmatrix} = \begin{bmatrix} (X_1X_1')^{-1}X_1 \\ (X_2X_2')^{-1}X_2(I - X_1'(X_1X_2')^{-1}X_1) \end{bmatrix} Y' \tag{2-6}$$

$\hat{B}_1$와 $\hat{B}_2$는 B_1과 B_2의 불편추정량이나 $\overline{B}_1$와 $\overline{B}_2$는 B_1과 B_2의 불편추정량이 아니며 편기(bias)가 존재한다.

(1-3)에서

$$X_1Y' = X_1X_1'\hat{B}_1' + X_1X_2'\hat{B}_2$$

이므로

$$\overline{B}_1' = (X_1X_1')^{-1}X_1Y'$$
$$= (X_1X_1')^{-1}(X_1X_1'\hat{B}_1' + X_1X_2'\hat{B}_2')$$
$$= \hat{B}_1' + (X_1X_1')^{-1}X_1X_2'\hat{B}_2'$$
$$\overline{B}_2' = (X_2X_2')^{-1}X_2(I - X_1'(X_1X_1')^{-1}X_1)Y'$$

즉,

$$(X_1X_2')^{-1}\overline{B}_2' = X_2Y' - X_2X_1'(X_1X_1')^{-1}X_1Y'$$
$$= (X_2X_1'\hat{B}_1' + X_2X_2'\hat{B}_2') - X_2X_1'(X_1X_1')^{-1}(X_1X_1'\hat{B}_1' + X_1X_2'\hat{B}_2')$$
$$= (X_2X_2' - X_2X_1'(X_1X_1')^{-1}X_1X_2')\hat{B}_2' \tag{2-8}$$

그러므로 편차항은 명백히

$$\overline{B}_1' - E(\overline{B}_1') = B_1' - (B_1' + (X_1X_1')^{-1}X_1X_2'B_2')$$
$$= -(X_1X_1')^{-1}X_1X_2'B_2' \tag{2-9}$$
$$\overline{B}_2' - E(\overline{B}_2') = B_2' - (B_2' - (X_2X_2')^{-1}X_2X_1'(X_1X_1')^{-1}X_1X_2'B_2')$$
$$= (X_2X_2')^{-1}X_2X_1'(X_1X_1')^{-1}X_1X_2'B_2' \tag{2-10}$$

가 된다.

參 考 文 獻

〔1〕 Freund, Rudolf J., Vail, Richard W. and Clunies, Ross C. W., "Residual Analysis," *Journal of the American Statistical Association*, Vol. 56, (1961), pp.98~104.

[2] Goldberger, Arthur S, and Jochems, D.B., "Note on Stepwise Least Squares," *Journal of the American Statistical Association*, Vol. 56, (1961), pp.107.

[3] Goldberger, Arthur S., "Stepwise Least Squares: Residual Analysis and Specification Error," *Journal of the American Statistical Association*, Vol. 56, (1961), pp.998~1000.

[4] Kabe, D. G., "Stepwise Multivariate Linear Regression," *Journal of the American Statistical Association*, Vol. 58, (1963), pp.770~771.

[5] Searle, S. R. and Hausman W.H., *Matrix Algebra for Business and Economics*, Wiley-Interscience/A Division of John Wiley and Sons, New York, London. Sydney, Toronto, 1970.

[6] Johnston J., *Econometric Methods*, 2nd ed., McGraw Hill Kogakusha, Ltd., 1972.

<ABSTRACT>

Alternative Derivation of Stepwise Multivariate Linear Regression

M. W. Shin
J. S. Kim

Freund, Vail, and Ross[1], Goldberger and Jochems [2] and Goldberger [3] have given some results for the stepwise estimation of the parameters of a univariate regression model, D.G. Kabe gave similar results for a multivariate linear regression model. We give here alternative derivation of some results derived by D.G Kabe.